Defense Threat Reduction Agency
8725 John J. Kingman Road, MS
6201 Fort Belvoir, VA 22060-6201

DTRA-TR-16-83

TECHNICAL REPORT

Cascading Failures in Networks: Inference, Intervention and Robustness to WMDs

Distribution Statement A. Approved for public release; distribution is unlimited.

August 2016

HDTRA1-13-1-0024

Sujay Sanghavi

Prepared by:
University of Texas, Austin
110 Inner Campus Drive
Austin, TX 78705

DESTRUCTION NOTICE:

Destroy this report when it is no longer needed.
Do not return to sender.

PLEASE NOTIFY THE DEFENSE THREAT REDUCTION AGENCY, ATTN: DTRIAC/ J9STT, 8725 JOHN J. KINGMAN ROAD, MS-6201, FT BELVOIR, VA 22060-6201, IF YOUR ADDRESS IS INCORRECT, IF YOU WISH IT DELETED FROM THE DISTRIBUTION LIST, OR IF THE ADDRESSEE IS NO LONGER EMPLOYED BY YOUR ORGANIZATION.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*
00-08-2016

2. REPORT TYPE
Technical

3. DATES COVERED *(From - To)*
February 1, 2013 - February 1, 2015

4. TITLE AND SUBTITLE
Cascading Failures in Networks: Inference, Intervention and Robustness to WMDs

5a. CONTRACT NUMBER

5b. GRANT NUMBER
HDTRA1-13-1-0024

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Sujay Sanghavi

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Texas, Austin
110 Inner Capus Drive
Austin, TX 78705

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Defense Threat Reduction Agency
8725 John J. Kingman Road, STOP 6201
Fort Belvoir, VA 22060-6201

10. SPONSOR/MONITOR'S ACRONYM(S)
DTRA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
DTRA-TR-16-83

12. DISTRIBUTION / AVAILABILITY STATEMENT
Distribution Statement A. Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Modern networks are prone to cascading failures: events where the initial destruction/compromising of a few nodes results in the successive and snowballing failure of a large portion of the network. WMD attacks are particularly likely to set off pernicious cascades, because they affect several interacting network (power, communication, human) simultaneously, and because the initially affected set of nodes itself is sizable. This proposal aims to significantly further our current limited and as-hoc understanding of cascading failures in networks.

15. SUBJECT TERMS
WMD; power grid; networks; data

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	SAR	12	Paul Tandy
Unclassified	Unclassified	Unclassified			**19b. TELEPHONE NUMBER** *(include area code)* 703-767-4663

UNIT CONVERSION TABLE

U.S. customary units to and from international units of measurement*

U.S. Customary Units	Multiply by → / ← Divide by†		International Units
Length/Area/Volume			
inch (in)	2.54	$\times 10^{-2}$	meter (m)
foot (ft)	3.048	$\times 10^{-1}$	meter (m)
yard (yd)	9.144	$\times 10^{-1}$	meter (m)
mile (mi, international)	1.609 344	$\times 10^{3}$	meter (m)
mile (nmi, nautical, U.S.)	1.852	$\times 10^{3}$	meter (m)
barn (b)	1	$\times 10^{-28}$	square meter (m^2)
gallon (gal, U.S. liquid)	3.785 412	$\times 10^{-3}$	cubic meter (m^3)
cubic foot (ft^3)	2.831 685	$\times 10^{-2}$	cubic meter (m^3)
Mass/Density			
pound (lb)	4.535 924	$\times 10^{-1}$	kilogram (kg)
unified atomic mass unit (amu)	1.660 539	$\times 10^{-27}$	kilogram (kg)
pound-mass per cubic foot (lb ft^{-3})	1.601 846	$\times 10^{1}$	kilogram per cubic meter (kg m^{-3})
pound-force (lbf avoirdupois)	4.448 222		newton (N)
Energy/Work/Power			
electron volt (eV)	1.602 177	$\times 10^{-19}$	joule (J)
erg	1	$\times 10^{-7}$	joule (J)
kiloton (kt) (TNT equivalent)	4.184	$\times 10^{12}$	joule (J)
British thermal unit (Btu) (thermochemical)	1.054 350	$\times 10^{3}$	joule (J)
foot-pound-force (ft lbf)	1.355 818		joule (J)
calorie (cal) (thermochemical)	4.184		joule (J)
Pressure			
atmosphere (atm)	1.013 250	$\times 10^{5}$	pascal (Pa)
pound force per square inch (psi)	6.984 757	$\times 10^{3}$	pascal (Pa)
Temperature			
degree Fahrenheit (°F)	$[T(^{o}F) - 32]/1.8$		degree Celsius (°C)
degree Fahrenheit (°F)	$[T(^{o}F) + 459.67]/1.8$		kelvin (K)
Radiation			
curie (Ci) [activity of radionuclides]	3.7	$\times 10^{10}$	per second (s^{-1}) [becquerel (Bq)]
roentgen (R) [air exposure]	2.579 760	$\times 10^{-4}$	coulomb per kilogram (C kg^{-1})
rad [absorbed dose]	1	$\times 10^{-2}$	joule per kilogram (J kg^{-1}) [gray (Gy)]
rem [equivalent and effective dose]	1	$\times 10^{-2}$	joule per kilogram (J kg^{-1}) [sievert (Sv)]

*Specific details regarding the implementation of SI units may be viewed at http://www.bipm.org/en/si/.
†Multiply the U.S. customary unit by the factor to get the international unit. Divide the international unit by the factor to get the U.S. customary unit.

Please answer all sections of the document. You are welcome to use figures and tables to complement or enhance the text. For annual reports, please only describe work for the period of performance (September 1, 2012 - June 30, 2013). For final reports, please describe the comprehensive effort.

Grant/Award #: HDTRA1-13-1-0024
PI Name: Sujay Sanghavi
Organization/Institution: University of Texas, Austin
Project Title: Cascading Failures in Networks: Inference, Intervention and Robustness to WMDs

What are the major goals of the project?
List the major goals of the project as stated in the approved application or as approved by the agency. If the application lists milestones/target dates for important activities or phases of the project, identify these dates and show actual completion dates or the percentage of completion. Generally, the goals will not change from one reporting period to the next. However, if the awarding agency approved changes to the goals during the reporting period, list the revised goals and objectives. Also explain any significant changes in approach or methods from the agency approved application or plan.

WMD attacks on modern networks are prone to creating cascading failures: events where the initial destruction/compromising of a few nodes results in the successive and snowballing failure of a large portion of the network.

Several examples of such outcomes come from infrastructure networks: the power grid is famously prone to cascade, as illustrated by mass blackouts caused by relatively small trigger events. Another example is transportation networks, where disruption in one part can lead to delays, cancellations etc. A third example is the Internet, where targeted disruption of a few key ISPs could lead to loss of connectivity to large parts of the web. Other examples of cascades come from diseases (indeed the classic models for cascades are often termed "epidemic models" for this reason).

Instead of narrowly focusing on specific settings of cascades, this proposal takes a broader view and aims to develop fundamental new mathematical understanding and algorithmic tools for learning and combating cascades. In particular, it aims to significantly further our current limited and ad-hoc understanding of cascading failures in networks, from three angles:

(i) Inference of key network structure and vulnerabilities from past events. In particular, the classic approach to the study of cascades (or epidemic processes) is to start with a model for the network and the spreading statistics, and derive typical cascade patterns. This task turns that on its head, and tries to learn the most representative and predictive network and model given past cascade events.

(ii) Intervention via rapid detection of an unfolding cascade, and response in the form of quarantining. In particular, once a good predictive model of the network and spread process is at hand, this task involves pre-computing best response strategies based both on networks structure and on preliminary data from the actual spread.

(iii) Developing incentives for better design, and forensics to trace the source/trajectory in cascade aftermaths. We draw on techniques from high-dimensional statistical machine learning, convex optimization, combinatorics and auction mechanism design for networks.

What was accomplished under these goals?
For this reporting period describe: 1) major activities; 2) specific objectives; 3) significant results, including major findings, developments, or conclusions (both positive and negative); and 4) key outcomes or other achievements. Include a discussion of stated goals not met. As the project progresses, the emphasis in reporting in this section should shift from reporting activities to reporting accomplishments.

1) Major Activities

Research:

This project is focused on taking a fundamental and data analytical approach to understanding cascading failures in networks, in the context of rare but large-scale disruptions like WMD attacks. Specifically, while the study of cascades has classically been “model driven” (where a fixed network graph and spread model is posited, and different cascade eventualities are investigated), this proposal aimed to focus on the inverse problem and build a “data driven” approach (where we observe past related cascade events to learn the network structure and spread model that best explains it).

Towards this end, we have several significant research thrusts:

(1) *Graph clustering:* This is a classic and fundamental problem: given a graph, partition nodes into groups so that the density of connectivity within groups is more than the density across groups. We made advances on two fronts: on the one hand, we characterized outer bounds - the best performance that any (even possibly unrealistic) algorithm can ever hope to achieve. This serves as a universal benchmark against which to compare the performance of every other (more computationally feasible) method. This work was awarded a **best paper award** at the Conference on Learning Theory (COLT), one of the most prestigious venues for machine learning. On the other hand, we also developed a way to recast graph clustering as convex optimization, enabling the use of the vast quiver of methods for continuous optimization to solve the inherently discrete clustering problem. This work has led to two conference papers (in NIPS and ICML) and one journal paper (in JMLR); taken together these papers have over 200 citations in the two years since their publication.

(2) Classifying cascades: Once the graph is ascertained, the next step is to find the underlying cause(s) for each one. In particular, cascade events can happen, and spread significantly, due to several underlying factors. Again, we aim to take a data-driven approach and find these underlying causes from observed cascade spreads. In machine learning terms, this can be cast as a latent variable problem: each underlying cause is a particular configuration of an a-priori unknown variable, and one needs to find both the variable values and the statistical relation between these values and observed cascade patterns. The statistical complexity of the cascade context presents several challenges; in this thrust we developed a new way to leverage (relatively very little) human input to efficiently and effectively analyze large-scale network events.

Human resource development:

This grant has allowed us to attract top-level talent to UT, and to nurture them to greater heights. The postdoc hired on this project, **Joe Neeman, will join the faculty of Mathematics at UT Austin as an Assistant Professor (tenure track)**. Joe also had faculty offers from Yale University, Umiversity of Michigan and Cornell. This project has been instrumental in supporting the research that got him hired. Joe previously had a PhD from UC Berkeley, where he did some highly regarded work on random graph theory.

This project has also part-supported a graduate student, **Praneeth Netrapalli**. Praneeth recently joined a very prestigious position at **Microsoft Research (MSR) as a full-time researcher**; before this he held a very coveted postdoc position at MSR as well. Again, this project enabled Praneeth to get some very widely cited results.

2) Specific Objectives

This grant aimed to develop the basic theory and algorithms for an "inverse problem" or "data-driven" study of cascades – specifically, learning about how they start, spread and can possibly be contained, not by positing a model but by learning from past cascade observations.

Towards this end, we focused on two specific objectives: finding clusters of co-dependent nodes that would be susceptible to and enable fast-spreading cascades, and finding the latent causes of cascades in large networks.

The "Research" subsection of Major Activities above gives further details on our approach and results for each objective.

3) Significant Results

The following papers were published/accepted in top venues in this year, as part of this project:

(1) "Improved graph clustering" – accepted to IEEE Transactions on Information Theory
(2) "Clustering partially observed graphs via convex optimization" – accepted to Journal of Machine Learning Research
(3) "Topic modeling from network spread" – accepted to SIGMETRICS 2014
(4) "Belief propagation, robust reconstruction and optimal recovery of block models" – accepted to COLT 2014 (best paper award)

Taken together, the above papers have over 250 citations in the approximately two years since their publication. All the venues – COLT, SIGMETRICS, JMLR, Trans. Information Theory – are the absolute top in their respective fields (machine learning, networks, information theory).

4) Key outcomes

Human resource development:

Joe Neeman, the postdoc funded on this project, is to join UT Austin as an Assistant Professor in Mathematics.
Praneeth Netrapalli, the student funded on this project, has graduated with a PhD and joined Microsoft Reserarch for a prestigious postdoc. Subsequently, he was offered – and has accepted – a full-time researcher position there. This is one of the most sought-after jobs in pure research, both in academia and industry.

Research:

The work funded through this project, although recent, has had impact in the community:

“Clustering sparse graphs,” a paper from the first year of this project, has already been cited 27 times; the total citations for our graph clustering work of the last year or so is over a 100.

“Learning the graph of epidemic cascades,” also a paper from the first year, has been cited 25 times in its the 1.5 years since it was published.

“Belief propagation, robust reconstruction and optimal recovery of block models” received the best paper award in the Conference on Learning Theory (COLT) 2014.

What opportunities for training and professional development has the project provided?
If the research is not intended to provide training and professional development opportunities or there is nothing significant to report during this reporting period, state "Nothing to Report." Describe opportunities for training and professional development provided to anyone who worked on the project or anyone who was involved in the activities supported by the project. "Training" activities are those in which individuals with advanced professional skills and experience assist others in attaining greater proficiency. Training activities may include, for example, courses or one-on-one work with a mentor. "Professional development" activities result in increased knowledge or skill in one's area of expertise and may include workshops, conferences, seminars, study groups, and individual study. Include participation in conferences, workshops, and seminars not listed under major activities.

This project has had a huge positive impact on the development of talent in the mathematics of networks. Specifically:

1. The postdoc hired on this project, **Joe Neeman, will join the faculty of Mathematics at UT Austin as an Assistant Professor (tenure track)**. This project has been instrumental in supporting the research that got him hired. Joe previously had a PhD from UC Berkeley.

2. This project has also part-supported a graduate student, **Praneeth Netrapalli**, who has since graduated with a PhD. Immediately after his PhD, Praneeth joined Microsoft Research for a very prestigious and competitive postdoc (very few are awarded across the entire country). Praneeth has since followed this up with a permanent researcher position (again a very coveted job) in Microsoft Research.

How have the results been disseminated to communities of interest?
If there is nothing significant to report during this reporting period, state "Nothing to Report." Describe how the results have been disseminated to communities of interest. Include any outreach activities that have been undertaken to reach members of communities who are not usually aware of these research activities, for the purpose of enhancing public understanding and increasing interest in learning and careers in science, technology, and the humanities.

The project has resulted in the following publications in top venues:

(1) "Improved graph clustering" – accepted to IEEE Transactions on Information Theory

(2) "Clustering partially observed graphs via convex optimization" – accepted to Journal of Machine Learning Research

(3) "Topic modeling from network spread" – accepted to SIGMETRICS 2014

(4) "Belief propagation, robust reconstruction and optimal recovery of block models" – accepted to COLT 2014 (best paper award)

Additionally, the PI has given talks on this at high-profile invited venues in conferences (Allerton, ITA) and departmental colloquia (UCLA, MIT, Boston Univ.)

Finally, UT Austin has an active industry outreach program called WNCG; this involves both leaders (VPs etc) and technical staff from industry having visits to UT, and facilitating visits at their companies. Our work on cascades has found resonance in several contexts, e.g. loss of connectivity of cellphone networks under attack etc.

What do you plan to do during the next reporting period to accomplish the goals?
If there are no changes to the agency-approved application or plan for this effort, state "No Change." Describe briefly what you plan to do during the next reporting period to accomplish the goals and objectives.

Not applicable.

DEPARTMENT OF DEFENSE

DEFENSE THREAT REDUCTION
AGENCY
8725 JOHN J. KINGMAN ROAD
STOP 6201
FORT BELVOIR, VA 22060
ATTN: P. TANDY

DEFENSE TECHNICAL
INFORMATION CENTER
8725 JOHN J. KINGMAN ROAD,
SUITE 0944
FT. BELVOIR, VA 22060-6201
ATTN: DTIC/OCA

DEPARTMENT OF DEFENSE CONTRACTORS

QUANTERION SOLUTIONS, INC.
1680 TEXAS STREET, SE
KIRTLAND AFB, NM 87117-5669
ATTN: DTRIAC